

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

CSE Journal Articles

Computer Science and Engineering, Department of

2016

A Spatio-Temporal Analysis of Mitochondrial DNA Haplogroup I

Peter Revesz

University of Nebraska-Lincoln, prevesz1@unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/csearticles>

Revesz, Peter, "A Spatio-Temporal Analysis of Mitochondrial DNA Haplogroup I" (2016). *CSE Journal Articles*. 197.
<https://digitalcommons.unl.edu/csearticles/197>

This Article is brought to you for free and open access by the Computer Science and Engineering, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in CSE Journal Articles by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

A Spatio-Temporal Analysis of Mitochondrial DNA Haplogroup I

Peter Z. Revesz^{1,a}

¹ Department of Computer Science, University of Nebraska-Lincoln, Lincoln, NE, 68588, USA

Abstract. The recent recovery of ancient DNA from a growing number of human samples shows that mitochondrial DNA haplogroup I was introduced to Europe after the end of the Last Glacial Maximum. This paper provides a spatio-temporal analysis of the various subhaplogroups of mitochondrial DNA I. The study suggests that haplogroup I diversified into haplogroups I1, I2'3, I4 and I5 at specific regions in Eurasia and then spread southward to Crete and Egypt.

1 Introduction

Ancient mitochondrial DNA (mtDNA) provides important insights on the movement and spread of human populations out of Africa. In particular, European populations exhibit some remarkable changes after the end of the Last Glacial Maximum. The changes in the early postglacial period are thought to be the result of the arrival of new human population groups to Europe. These new populations brought to Europe the I and W mitochondrial DNAs [6]. The origin and exact spread of these two mitochondrial DNAs is little understood. This paper provides a spatio-temporal analysis of mtDNA I.

This paper is organized as follows. Section 2 presents some related work. Section 3 describes the source data and our data curating to refine and improve the haplogroup classifications and find the precise latitude and longitude coordinates of the location where each sample was discovered. Section 4 describes our spatio-temporal analysis of the curated data. Finally Section 5 gives some conclusions and directions for future work.

2 Related work

Several papers have attempted to trace the movement and spread of human populations based on ancient DNA data. Allentoft et al. [1] considers Eurasian populations and traces of Indo-European migrations in the Bronze Age. Other papers have made DNA-based connections between past and present day populations, such as Brandt et al. [3] between ancient DNA and present Central Europeans and Vai et al. [14] between Medieval Lombards and present day Italians.

The difficulty of making connections between pairs of populations lies in the fact that most human populations are genetically fairly mixed, making the comparison of populations rely on complex statistical similarities according to several proposals including a

recent one by the author [11]. For example, modern European human populations in every region contain a large number of mitochondrial DNA haplogroups. In contrast, Paleolithic human samples contain only a limited number of mtDNA haplogroups, mostly belonging to U. The genetics of European Neolithic human populations is relatively well-studied and traced back to Anatolia and the Near East [2], [4], [5]. However, mtDNA haplogroup I is associated neither with the Paleolithic nor with the Neolithic expansion. Hervella et al. [6] show that mtDNA haplogroup I occurs in Paternanbidea, Spain c. between 6000 BC. This early presence in Spain suggests that mtDNA I expanded to Eurasia from the Franco-Cantabrian glacial refugium after the Last Glacial Maximum. However, Olivieri et al. [8] argue that mtDNA haplogroup I originates from the Near East. That discrepancy and the uncertainty regarding the origin of the subhaplogroups of haplogroup I motivates the present study.

3 Source data and data curation

Several websites are specialized to record ancient mitochondrial DNA samples. Below we used the ancestral journals online database and website, which is available at the following Internet site: <http://www.ancestraljourneys.org/ancientdna.shtml>.

Table 1 lists from thirteen cultures fourteen samples that can be classified into one of the mtDNA haplogroups I1, I2, I3, I4 and I5. The Unetice culture had both I1 and I3 samples. For each of the sample locations, we found its latitude and longitude using Google map.

Some of the samples were classified as simply mtDNA I, but we could use the latest version of the mtDNA haplogroup classification table (Build 17, February 18, 2016) at www.PhyloTree.org to further refine the classification to I1, I2, I3, I4 or I5 haplogroup.

^a Corresponding author: revesz@cse.unl.edu

Table 1. Cultures that contain mitochondrial DNA I1, I2, I3, I4 or I5 samples.

Type	Culture	Date	Place	Latitude	Longitude
Neolithic	Cami de Can Grau Granollers	3500-3000 BC	Cami de Can Grau Granollers, Barcelona, Spain	41.43	2.12
Bronze	Yamnaya	3000-2500 BC	Kirovohrad, Ukraine	48.30	32.16
Bronze	Bell Beaker	2600-2500 BC	Kromsdorf, Germany	50.98	11.37
Bronze	Minoan	2200-1700 BC	Charalambos cave, Crete, Greece	35.18	25.44
Bronze	Unetice	2200-1550 BC	Eulau, Germany	51.17	11.85
Bronze	Unetice	2200-1550 BC	Esperstedt, Germany	51.42	11.67
Bronze	Mezhovskaya	1598-1398 BC	Kapova cave, Russia	53.04	57.06
Bronze	Karasuk	1416-1268 BC	Sabinka, near Karasuk, Russia	53.43	78.03
Bronze	Vodokhranilische	1400-1000 BC	Vodokhranilische, Kazakhstan	49.52	49.86
Iron	Norabak	1209-1009 BC	Norabak, Armenia	40.13	45.86
Iron	Scythian	500-200 BC	Rostov-on-Don, Russia	47.14	39.42
Iron	Egyptian	402-385 BC	Faiyum Valley, Egypt	29.57	30.66
Medieval	Lombard	570-650 AD	Collegno, Italy	45.50	7.35
Medieval	Hungarian	975-1000 AD	Sárrétudvar, Hungary	47.25	21.17



Figure 1. The distribution of mtDNA haplogroup samples I1 (blue), I2 and I3 (yellow), I4 (red) and I5 (green).

Table 2. Ancient DNA samples that belong to some branch of mtDNA haplogroup I. The samples and their classifications are from the website <http://www.ancestraljourneys.org/ancientdna.shtml>. Some of the classifications are updated. In case of updates, the original classification is on the left and the new classification is on the right of the > symbol,

CULTURE:	Cami de Can	Bell Beaker	Unetice	Yamnaya	Lombard	Egypt	Unetice	Scythian	Minoan	Hungarian	Mezhdovskaya	Vodokhranilische	Norabak	Karasuk
YEAR:	3500-3000	2600-2500	2200-1550	3000-2500	570-650 AD	402-385	2200-1550	500-200	2200-1700	975-1000 AD	1598-1398	1400-1000	1209-1009	1416-1268
GROUP:	I1c1	I1a1	I1a1	I > I1a	I2a	I2	I3a	I3 > I3a	I5a	I > I5a	I5c > I5a2+	I > I5a4	I5c	I4a1
T10034C I														
G16129A I	x	x		x		x		x	x		x	x	x	x
455.1T I1														
G6734A I1														
G9966A I1														
T16311C I1		x		x										
A3447G I1a														
G8616t I1a														
T16172C I1a				x										
G203A I1a1				x										
C16264T I1c	x													
G16319A I1c	x													
T16362C I1c	x													
T9386C I1c1														
C16270T I1c1	x													
T152C I2'3						x					x			
G207A I2'3						x								
A15758G I2						x								
G16145A I2a					x									
T239C I3														
T16086C I3a							x	x			x			
C16494T I3b														
T2628C I3c														
T8260C I3d				x										
G8519A I4														x
A10819G I4a														x
T16304C I4a1														x
C9851T I4a2														
A2308G I4b														
G4113A I4b														
C13662T I4b														
A14233G I5													x	
T5074C I5a									x					
C16148T I5a										x				
T16086C I5a2+							x	x			x			
C16294T I5a4												x		
G709A I5b														
G6506T I5b											x			
C16169T I5c													x	

In particular, we improved the classification of one Yamnaya sample from I to I1a, one Vodokhranilische sample from I to I5a4, and one Hungarian sample from I to I5a. In addition, we could refine the classification of a Scythian I3 sample to I3a. Table 2 gives a detailed explanation of the reason for these refinements. As Table 2 shows, there are still some ambiguities regarding the classification of a few samples. For example, the Mezhevskaya sample can be considered either I5c or I5a2 as shown in Table 2. The ancestral journeys online database contains some other mtDNA I samples that could not be further refined, including currently the earliest mtDNA haplogroup I sample from Paternabidea, Spain from the Neolithic [6].

In Table 2 the first four data columns contain I1 samples, the next four contain I2 and I3 samples, the next five contain I5 samples, and the last contains a single I4 sample. The I2 and I3 haplogroups have a common ancestor called mtDNA I2'3. Therefore, in this paper we consider them as one group. We color-coded the haplogroups as blue for I1, golden for I2 and I3, green for I5, and red for I4. Using a similar color scheme, Figure 1 shows the location of each sample on a map.

4 Spatio-temporal analysis

At first we computed the centroid of each haplogroup. The result is shown in Table 3.

Table 3. The haplogroup centroid values.

Haplogroup	Latitude	Longitude
I1	47.97	14.375
I2'3	43.4075	22.275
I4	53.43	78.03
I5	45.024	39.878

The data in Table 3 needs to be considered together with the map in Figure 1. In combination, they allow us to draw some conclusions regarding the origin of the various haplogroups in this study. In particular the map suggest the following:

I1: This haplogroup is more western (with longitude 14.375) than other haplogroups and also more northern (47.97) than other haplogroups except for I4. Hence it is likely that this haplogroup originated in northwestern Europe.

I2'3: This haplogroup occurs in an old Unetice sample and three considerably more recent samples near the shores of the Mediterranean and Black Seas. Hence it is likely that this group also originated in northern Eurasia and the three later samples are the result of population movements facilitated by the seas.

I4: This haplogroup occurs only in one sample, which happens to be the easternmost of all the samples. Hence it is likely that this haplogroup originates somewhere in central Eurasia.

I5: This haplogroup seems to be concentrated between and above the Black and the Caspian Seas. Hence that area is the likely origin of haplogroup I5. The presence of haplogroup I5 in some Minoan samples may be explained as a population movement from the Black Sea to Crete. The Hungarian sample may be explained as another population movement from the region north of the Black Sea to present day Hungary.

Overall, the ancient mtDNA data suggests that the origin of the I1, I2'3, I4 and I5 haplogroups are all connected with various regions in Europe and western Asia.

5 Conclusions and future work

The spatio-temporal analysis of this paper can be extended in many ways using the methods described in [9]. In particular, we are interested in modeling the spread of human populations as spatio-temporal objects that are described using constraint databases [7]. This may be further combined with an extension of phylogenetic tree algorithms, such as the *Common Mutations Similarity Matrix* algorithm [10] and the *Incremental Phylogenetics by Repeated Insertions* algorithm [12]. Finally, the rate of evolution of the various subhaplogroups of mtDNA I is another interesting topic to be further explored perhaps by applying techniques similar to those in the study of bacterial evolution [13].

The extended study could also benefit from other ancient mtDNA I samples that may become available in the near future. The study of ancient human DNA is an exciting area of research, and in that study mtDNA haplogroup I plays an interesting role as a testimony to the earliest meeting of Paleolithic survivors in glacial refugia and newcomers to the European continent from either North Africa or the Near East.

References

1. M. Allentoft et al., *Nature*, **522**, 167-172 (11 June 2015)
2. R. Bollongino, O. Nehlich, M.P. Richards, J. Orschiedt, M.G. Thomas, C. Sell, Z. Fajkošová, A. Powell, J. Burger, *Science*, **342**, 6157, 479-481 (25 Oct. 2013)
3. G. Brandt et al., *Science*, **342**, 6155, 257-261 (2013).
4. P. Brotherton, *Nature Comm.*, **4**, Art. No. 1764 (2013)
5. W. Haak, et al., *PloS Bio.*, **8**, 11, (November 2010)
6. M. Hervella et al., *Arch. and Anthropol. Sciences*, **8**, 1, 97-106 (March 2016)
7. P.C. Kanellakis, G.M. Kuper, P.Z. Revesz, J. of Comp. and Sys. Sciences, **51**, 1, 26-52 (1995)
8. A. Olivieri, M. Pala, F. Gandini, B.H. Kashani, U.A. Perego, S.R. Woodward, V. Grugni, V. Battaglia, O. Semino, A. Achilli, M.B. Richards, A. Torroni, *PLoS ONE*, **8**, 7, (2013)

9. P.Z. Revesz, *Introduction to Databases: From Biological to Spatio-Temporal*, (Springer, New York, 2010)
10. P.Z. Revesz, Proc. 4th ACM Int. Conf. on Bioinfo. and Comp. Bio., 731-734, (ACM Press, New York, USA, 2013)
11. P.Z. Revesz, Int. J. of Bio. and Biomed. Eng., **10**, 1, 124-133 (2016)
12. P.Z. Revesz and Z. Li, Int. J. of Bio. and Biomed. Eng., **10**, 1, 148-158 (2016)
13. M. Shortridge, T. Triplet, P.Z. Revesz, M. Griep, R. Powers, Comp. Bio. and Chem., **35**, 1, 24-33, (2011)
14. S. Vai et al., PLoS ONE **10**, 1, (2015)